Epidemic spreading in correlated complex networks

Marián Boguñá[1] and Romualdo Pastor-Satorras[2]

[1]*Departament de Física Fonamental, Universitat de Barcelona, Av. Diagonal 647, 08028 Barcelona, Spain*
[2]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*

(Dated: February 1, 2008)

We study a dynamical model of epidemic spreading on complex networks in which there are explicit correlations among the node's connectivities. For the case of Markovian complex networks, showing only correlations between pairs of nodes, we find an epidemic threshold inversely proportional to the largest eigenvalue of the connectivity matrix that gives the average number of links that from a node with connectivity k go to nodes with connectivity k'. Numerical simulations on a correlated growing network model provide support for our conclusions.

PACS numbers: 89.75.-k, 87.23.Ge, 05.70.Ln

Statistical physics has witnessed in recent years a renewed interest in graph theory due to the discovery that many natural and artificial systems can be described in terms of complex networks, in which the nodes represent typical units, and the links represent the interactions between pairs of units [1, 2]. The term complex network has been coined to refer to networks that typically exhibit two distinct properties: (i) A *scale-free* (SF) connectivity distribution. If we define the connectivity distribution, $P(k)$, as the probability that a node is connected to k other nodes, then SF networks are characterized by a power-law behavior $P(k) \sim k^{-\gamma}$, where γ is a characteristic exponent. This property renders classical models of random graphs [3], described by an exponentially bounded connectivity, inappropriate for the description of many real networks. (ii) The *small-world* property [4], that is defined by an average path length—average distance between any pair of nodes—increasing very slowly (usually logarithmically) with the network size N.

Apart from the empirical characterization of real complex networks and the development of new models, accounting for the observed properties, the interest in this field has been also driven by the discovery of the profound and subtle effects that the connectivity has on the behavior of dynamical systems defined on top of complex networks. These effects are particularly interesting in the study of disease transmission in SF networks, relevant for the understanding of the spreading of computer viruses [5] and sexually transmitted diseases [6]. Indeed, it was first noted in Ref. [7] that in *uncorrelated* SF networks with a connectivity exponent $\gamma \leq 3$, epidemic processes do not posses an epidemic threshold below which diseases cannot produce a macroscopic epidemic outbreak or the development of an endemic state. This feature, observed in several epidemic models [7, 8, 9, 10, 11], is deeply rooted in the presence of very large connectivity fluctuations in infinite SF networks.

The study of epidemic spreading in uncorrelated complex networks (that is, in graphs in which the connectivity of any node is independent of the connectivity of its neighbors) has been proved to be extremely successful, providing, for instance, the first satisfactory explanation of the long-standing problem of the generalized low prevalence of computer viruses without assuming any global tuning of microscopic parameters [7, 12]. Nevertheless, it represents a first approximation to real networks, that neglects the possibility of correlations in the connectivity of the nodes. The existence and importance of connectivity correlations has been recently pointed out in the literature. In fact, it has become clear that these correlations are critical in the understanding of the hierarchical structure of the Internet [13, 14, 15]. On the other hand, some growing network models have been proposed [16], in which correlations are spontaneously generated and have important effects in the percolation transition.

In this paper we shall present a study of epidemic spreading in complex random networks in which there are explicit correlations among the node's connectivities. We will consider in particular the subset of undirected *Markovian* random networks, that are completely defined by their connectivity distribution $P(k)$ and the conditional probability $P(k' \,|\, k)$ that a node of connectivity k is connected to a node of connectivity k'. These two functions can have any form (including SF behavior), and are assumed to be normalized ($\sum_k P(k) = \sum_{k'} P(k' \,|\, k) = 1$) and restricted by the connectivity detailed balance condition

$$kP(k' \,|\, k)P(k) = k'P(k \,|\, k')P(k') \equiv \langle k \rangle \, P(k, k'), \quad (1)$$

where the symmetric function $(2 - \delta_{kk'})P(k, k')$ is the joint probability that two nodes of connectivity k and k' are connected. The Markovian nature of this class of networks implies that all higher order correlations can be expressed as a function of $P(k)$ and $P(k' \,|\, k)$, allowing an exact treatment of epidemic models at the mean-field (MF) level. It is worth noticing, however, that a more detailed description, in terms of a Langevin equation (to be reported elsewhere [17]), yields exactly the same results, confirming the accuracy of the MF description. In this framework, the topologically relevant magnitude is the connectivity matrix $C_{kk'} = kP(k' \,|\, k)$, that measures the

average number of links that go from a node with connectivity k to nodes with connectivity k'. We will show that the epidemic threshold is related to the largest eigenvalue of this matrix. Extensive numerical simulations on a random correlated network model [16] confirm the predictions of the present analysis. During the completion of this work we became aware of two recent preprints [18, 19] in which it is also highlighted the general role of correlations in spreading and percolation in complex networks.

In order to study the effects of connectivity correlations in epidemic spreading, we will focus in the standard susceptible-infected-susceptible (SIS) model [20]. All the results, however, can be easily extended to the more general susceptible-infected-removed-susceptible (SIRS) model [17]. In the SIS model each node in the network represents an individual, and each link represents a connection along which the infection can propagate. Susceptible (healthy) nodes become infected with probability ν if they are connected to one or more infected nodes. On the other hand, infected nodes recover spontaneously with probability δ. The ratio of this two rates defines an effective spreading rate $\lambda = \nu/\delta$ (without lack of generality, we set $\delta = 1$). For homogeneous networks, in which each node has more or less the same number of connections, $k \simeq \langle k \rangle$, a general result states the existence of a finite epidemic threshold, separating an infected (endemic) phase, with a finite average density of infected individuals, from a healthy phase, in which the infection dies out exponentially fast. In terms of the average density of infected individuals $\rho(t)$ (the prevalence) we can describe the SIS model in homogeneous networks at a MF level by the following rate equation [8]

$$\partial_t \rho(t) = -\rho(t) + \lambda \langle k \rangle \rho(t) [1 - \rho(t)]. \qquad (2)$$

In this equation we have neglected higher order terms, since we are interested in the onset of the endemic state, close to the point $\rho(t) \sim 0$. Also, we have neglected correlations among nodes. That is, the probability of infection of a new node—the second term in Eq. (2)—is proportional to the infection rate λ, to the probability that a node is healthy, $1 - \rho(t)$, and to the probability that a link in a healthy node points to an infected node. This last quantity, assuming the *homogeneous mixing hypothesis* [21], is approximated for homogeneous networks as $\langle k \rangle \rho(t)$, i.e. proportional to the average number of connection and to the density of infected individuals, and independent of the connectivity. From Eq. (2) it can be proved the existence of an epidemic threshold $\lambda_c = \langle k \rangle^{-1}$ [20], such that $\rho = 0$ if $\lambda < \lambda_c$, while $\rho \sim (\lambda - \lambda_c)$ if $\lambda \geq \lambda_c$. In this context, it is easy to recognize that the SIS model is a generalization of the contact process model, widely studied as the paradigmatic example of an absorbing-state phase transition to a unique absorbing state [22].

For general complex networks, in which large connectivity fluctuations and correlations might be allowed, we must relax the homogeneous hypothesis made in writing Eq. (2) and work instead with the relative density $\rho_k(t)$ of infected nodes with given connectivity k; i.e. the probability that a node with k links is infected. Following Refs. [7, 8], the rate equation for $\rho_k(t)$ can be written as

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t) + \lambda k [1 - \rho_k(t)] \Theta_k(t). \qquad (3)$$

In this case, the creation term is proportional to the spreading rate λ, the density of healthy sites $1 - \rho_k(t)$, the connectivity k, and the variable $\Theta_k(t)$, that stands for the probability that a link emanating from a node of connectivity k points to an infected site. In the case of an uncorrelated random network, considered in Refs. [7, 8], the probability that a link points to a node of connectivity k' is *independent* of the connectivity k of the node from which the link is emanating. Therefore, in this case $\Theta_k = \Theta^{\rm nc}$ is independent of k and can be written as

$$\Theta^{\rm nc} = \frac{1}{\langle k \rangle} \sum_{k'} k' P(k') \rho_{k'}(t), \qquad (4)$$

since the probability that a node is pointing to a node of connectivity k' is proportional to $k'P(k')$. Substituting the expression (4) into Eq. (3), one can solve for the steady state solution and find the existence of an epidemic threshold λ_c, below which there are no solutions with a nonzero value of $\Theta^{\rm nc}$. The expression of the epidemic threshold for uncorrelated random networks is

$$\lambda_c^{\rm nc} = \frac{\langle k \rangle}{\langle k^2 \rangle}. \qquad (5)$$

For infinite SF networks with $\gamma \leq 3$, we have $\langle k^2 \rangle = \infty$, and correspondingly $\lambda_c^{\rm nc} = 0$. Finally, from the solution of ρ_k, one can compute the total prevalence ρ using the relation $\rho = \sum_k P(k) \rho_k$.

For a general network in which the connectivities of the nodes are correlated, the above formalism is not correct, since we are not considering the effect of the connectivity k into the expression for Θ_k. This effect can be taken into account, however, for Markovian networks, whose correlations are completely defined by the conditional probability $P(k' \,|\, k)$. In this case, it is easy to realize that the correct factor Θ_k can be written as

$$\Theta_k(t) = \sum_{k'} P(k' \,|\, k) \rho_{k'}(t), \qquad (6)$$

that is, the probability that a link in a node of connectivity k is pointing to an infected node is proportional to the probability that any link points to a node with connectivity k', times the probability that this node is infected, $\rho_{k'}(t)$, averaged over all the nodes connected to the original node. Eqs. (3) and (6) define together

the MF equation describing the SIS model on Markovian complex networks,

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t) + \lambda k\,[1-\rho_k(t)] \sum_{k'} P(k'\,|\,k)\rho_{k'}(t). \quad (7)$$

The exact solution of Eq. (7) can be difficult to find, depending on the particular form of $P(k'\,|\,k)$. However, it is possible to extract the value of the epidemic threshold by analyzing the stability of the steady-state solutions. Of course, the healthy state $\rho_k = 0$ is one solution. For small ρ_k, we can linearize Eq. (7), getting

$$\frac{d\rho_k(t)}{dt} \simeq \sum_{k'} L_{kk'}\rho_{k'}(t). \quad (8)$$

In the previous equation we have defined the Jacobian matrix $\mathbf{L} = \{L_{kk'}\}$ by

$$L_{kk'} = -\delta_{kk'} + \lambda k P(k'\,|\,k), \quad (9)$$

where $\delta_{kk'}$ is the Kronecker delta function. The solution $\rho_k = 0$ will be unstable if there exists at least one positive eigenvalue of the Jacobian matrix $\mathbf{L}$. Let us consider the *connectivity matrix* $\mathbf{C}$, defined by $C_{kk'} = kP(k'\,|\,k)$. Using the symmetry condition Eq. (1), it is easy to check that if v_k is an eigenvector of $\mathbf{C}$, with eigenvalue Λ, then $P(k)v_k$ is an eigenvector of the transposed matrix $\mathbf{C}^T$ with the same eigenvalue. From here it follows immediately that all the eigenvalues of $\mathbf{C}$ are real. Let Λ_m be the largest eigenvalue of $\mathbf{C}$. Then, the origin will be unstable whenever $-1 + \lambda\Lambda_m > 0$, which defines an epidemic threshold

$$\lambda_c = \frac{1}{\Lambda_m}, \quad (10)$$

above which the solution $\rho_k = 0$ is unstable, and another nonzero solution takes over as the actual steady-state—the endemic state.

It is instructive to see how this general formalism recovers previous results [7, 8], implicitly obtained for random uncorrelated networks. For any random network, in which there are no correlations among the connectivities of the nodes, we have that the connectivity matrix is given by $C^{\rm nc}_{kk'} = kP(k'/k) \equiv kk'P(k')/\langle k\rangle$, since the probability that a link points to a node of connectivity k' is proportional to $k'P(k')$. It is easy to check that the matrix $\{C^{\rm nc}_{k'k}\}$ has unique eigenvalue $\Lambda^{\rm nc}_m = \langle k^2\rangle / \langle k\rangle$, corresponding to the eigenvector $v^{\rm nc}_k = k$, from where we recover the now established result Eq. (5).

In order to check the theoretical prediction Eq. (10), we have performed numerical simulations of the network model proposed by Callaway *et al.* [16]. This model is very simply defined: each time step we add a new node, and, with probability δ, two nodes are randomly selected and joined with a link. This recipe yields a random network with a exponential connectivity distribution, $P(k) = (2\delta)^k/(1+2\delta)^{k+1}$, which shows correlations among the connectivities of the nodes. These correlations can be analytically computed by means of the joint probability $P(k,k')$, which fulfills the recursion relation [16]

$$P(k,k') = \frac{2\delta}{1+4\delta}[P(k-1,k') + P(k,k'-1)] + \frac{P(k-1)P(k'-1)}{1+4\delta}.$$

Even though it is not possible to solve for $P(k,k')$ in a closed form, we can numerically estimate the largest eigenvalue of the connectivity matrix $\mathbf{C}$ by generating a finite matrix $P(k,k')$ and using the relation (1) to find $P(k'\,|\,k)$. The correctedness of this numerical estimate is ensured by the exponential decay of $P(k,k')$ for large values of k and k', as can be seen from the recursion relation. The estimate of the largest eigenvalue of the connectivity matrix for Callaway's model is $\Lambda_m \approx 6.47656$, which yields an epidemic threshold $\lambda_c \approx 0.15$. This value is to be compared with the prediction for an uncorrelated network, $\lambda^{\rm nc}_c = \langle k\rangle / \langle k^2\rangle = 0.20$.

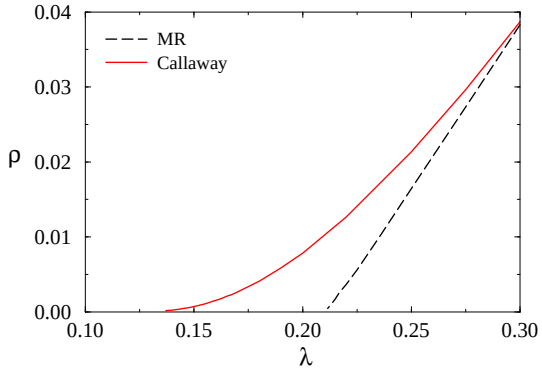

FIG. 1: Total prevalence ρ for the SIS model in Callaway's network and in a MR uncorrelated network with the same connectivity distribution.

Figure 1 shows the results of numerical simulations of the SIS model on Callaway's network, as well as on random networks with the same connectivity distribution, generated using the Molloy and Reed (MR) algorithm [23, 24]. The MR algorithm generates a random network with a prescribed connectivity distribution and no correlations among nodes, and thus it is expected to yield an epidemic threshold given by $\lambda^{\rm nc}_c$. Simulations were performed for a fixed value $\delta = 1$ in networks of size up to $N = 10^7$, averaging over at least 100 different starting configurations, performed on at least 10 different realizations of the network. Figure 1 depicts the steady state prevalence as a function of the spreading rate λ. For the MR network, the function $\rho(\lambda)$ shows a clear linear behavior. The epidemic threshold estimated from a least-squares fitting is $\lambda_c = 0.21 \pm 0.01$, in excellent agreement with the prediction for an uncorrelated network. On the other hand, Callaway's network exhibits a

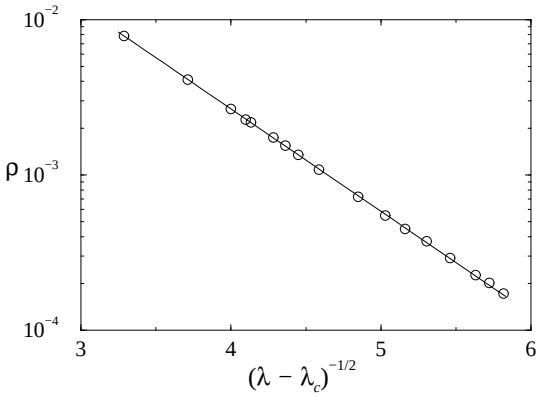

FIG. 2: Fit of the prevalence for the SIS model in Callaway's network to the form $\rho(\lambda) \sim \exp[-\alpha(\lambda - \lambda_c)^{-1/2}]$. The fit yields the values $\alpha = 1.52 \pm 0.02$ and $\lambda_c = 0.11 \pm 0.02$.

very different behavior, which might be compatible with the presence of a transition of infinite order [22]. In fact, Fig. 1 is reminiscent of the behavior found in Ref. [16] for the giant component of the network as a function of the parameter δ. In that work, the size of the giant component was fitted to an stretched exponential with exponent 1/2. Guided by this intuition, in Fig. 2 we perform a fit of the stationary prevalence in Callaway's model to the form $\rho(\lambda) \sim \exp[-\alpha(\lambda - \lambda_c)^{-1/2}]$ [16]. The fit yields a prefactor $\alpha = 1.52 \pm 0.02$, and an epidemic threshold $\lambda_c = 0.11 \pm 0.02$, smaller by a factor 2 than the value corresponding to an uncorrelated random network, and in quite good agreement with the prediction from the largest eigenvalue of the connectivity matrix.

In summary, we have shown that, in the presence of correlations, the epidemic threshold in complex networks is determined by the connectivity matrix $\mathbf{C}$, and not by the connectivity distribution $P(k)$, as happens in uncorrelated networks. This fact implies that the previously predicted null epidemic threshold for SF networks with $\gamma \leq 3$ might be shifted in correlated graphs, attaining a positive value depending on the nature of the correlations as given by the connectivity matrix. At this respect, it might be surprising that some complex networks, such as the Barabási-Albert (BA) graph [25], are exactly described at the uncorrelated level given by Eq. (3) with Θ_k independent of k [7, 8]. This fact must be taken as an evidence of the lack of correlations in the BA model; lack that, on the other hand, has been already checked numerically in Ref. [14]. The formalism presented in this paper represents a refinement over previous works because it includes the effects of correlations between pairs of nodes and, in this sense, it is exact for Markovian networks. Real networks, such as the Internet, however, posses a more complex correlation structure. Our formalism will provide an improved approximation to epidemic dynamics in these cases, but it still remains the task of ascertaining the effects of higher order correlations. Further work is necessary in this direction.

This work has been partially supported by the European Commission - Fet Open project COSIN IST-2001-33555. R.P.-S. acknowledges financial support from the Ministerio de Ciencia y Tecnología (Spain). We thank A. Vespignani for helpful comments and discussions.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[3] P. Erdös and P. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
[4] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
[5] R. Pastor-Satorras and A. Vespignani, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Berlin, 2002), e-print cond-mat/0205260.
[6] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, Nature **411**, 907 (2001).
[7] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[8] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).
[9] R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
[10] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **26**, 521 (2002).
[11] M. E. J. Newman, *Exact solutions of epidemic models on networks* (2001), e-print cond-mat/0201433.
[12] S. R. White, *Open problems in computer virus research*, Virus Bulletin Conference, Munich (1998), (http://www.av.ibm.com/ScientificPapers/White/Problems/Problems.html).
[13] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[14] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, *Large-scale topological and dynamical properties of Internet*, e-print cond-mat/0112400.
[15] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett **88**, 108701 (2002).
[16] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, Phys. Rev. E **64**, 041902 (2001).
[17] M. Boguñá and R. Pastor-Satorras, (in preparation).
[18] M. E. J. Newman, *Assortative mixing in networks*, e-print cond-mat/0205405.
[19] V. M. Eguíluz and K. Klemm, *Epidemic threshold in structured scale-free networks*, e-print cond-mat/0205439.
[20] O. Diekmann and J. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation* (John Wiley & Sons, New York, 2000).
[21] R. M. Anderson and R. M. May, *Infectious diseases in humans* (Oxford University Press, Oxford, 1992).
[22] J. Marro and R. Dickman, *Nonequilibrium phase transitions in lattice models* (Cambridge University Press, Cambridge, 1999).
[23] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).

[24] M. Molloy and B. Reed, Combinatorics, Probab. Comput. **7**, 295 (1998).

[25] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).